U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 219

The Global Data Base in 1980

Ronald D. McPherson
Development Division

JULY 1980

I. Introduction

We have previously noted that one of the most important trends affecting operational meteorology over the past decade or so is the changing character of the global data base. This lecture discusses its current status. In the next section, the evolution of the operational data base over the last decade as seen at NMC is discussed in general terms. The third section presents a series of short essays on the main components of the data base, in terms of the number of reports from each source, their distribution and availability, and their characteristics and error structure. Finally, the lecture concludes with a brief description of automatic data processing procedures currently in use at NMC.

II. Evolution of the global data base

During the first decade of operational numerical weather prediction, the upper air data base at NMC consisted primarily of radiosonde reports from the Northern Hemisphere. These were mostly located over populous continents, although a few - generally less than 20* - were on board ships (Ocean Station Vessels) supported by the governments of several countries. These were positioned partly for meteorological reasons, and partly to assist in rescue operations in the event of marine disaster. In addition, there were a few aircraft observations, but these were in general regarded as unreliable by meteorological analysts. At the surface a few thousand weather stations transmitted synoptic reports. A few hundred of these reports were from ships at sea. By far the largest majority of all reports were from the Northern Hemisphere.

Although the combined surface and upper air data base was reasonably adequate for short-range predictions for the interiors of continents, long-range forecasts and predictions for areas immediately downstream from data-void areas suffered severe limitations. As a partial attempt to alleviate the problems in numerical weather prediction caused by lack of data over large areas, the practice of manual intervention, or "bogusing", was devised. Skilled analysts would perform subjective analyses based on all information available, including information not readily ingested by automated analysis techniques. Cloud imagery from orbiting satellites, available by the mid-1960's, is one example of such data. From the subjective analyses, the analyst would manufacture data - "bogus" reports - to supplement the available real data over oceanic areas. Such intervention is obviously time-consuming, and in an operational environment with the pressure of meeting transmission deadlines, it was of necessity limited to one or two isobaric levels.

Thus at the midpoint of the 1960's, the operational data base at NMC approached adequacy for synoptic-scale phenomena over populous land areas of the Northern Hemisphere. Subjectively-generated data at one or two levels, plus aircraft reports, served as the primary basis of

---

* Only in 1945 (43) and 1946(34) did the number of OSV's exceed 20. (A. Thomas, NMC; personal communication).

meteorological analysis over oceans. This permitted hemispheric predictions to be calculated out to two or three days with some skill. No attempt was made to extend numerical weather prediction to the globe.

The first important augmentation of the data base occurred in the late 1960's. A meteorological satellite equipped with a television camera, the Applications Technology Satellite (ATS-1), was launched into geosynchronous orbit on June 6, 1966. Since the satellite remained essentially fixed over one location on the earth, sequential images of cloud masses could be followed in time. Individual cloud "targets" could be tracked and their spatial displacements over the interval of the image sequence calculated. To the extent that the cloud targets move with the wind, the displacements can be converted to a vector wind at some altitude which must be estimated. By the end of the decade these vectors were being used operationally, although the date of their entry into the data base is lost in the mists of time.

A second major data source consisted of remotely-sensed temperature soundings from passive radiometers on board polar-orbiting satellites. These instruments sense upwelling radiation at distinct frequencies emitted by the atmosphere. The frequencies are chosen so that the emissivity at each frequency reaches its maximum at a different level in the atmosphere. Thus the radiation perceived by the radiometer at a given frequency originates mostly in an approximately-definable layer. Since the emissivity of any layer is a function of the emittor's temperature, measured radiation at several frequencies makes it possible to infer the mean temperature of several layers. Overlapping of the layers requires some unscrambling of the data to determine the independent temperature estimates, a process usually denoted "retrieval". Data from this source first entered the data base May 28, 1969 .

By the end of the second decade of numerical weather prediction, these two sources contributed significantly to the operational data base. In addition, the great expansion of commercial aviation produced a substantial increase in the number of aircraft reports. Table 1 illustrates the growth of the data base. The first column gives the average number of reports from different sources received at NMC by 10 hours after midnight or noon GMT for one week in November 1975. In the original source (McDonell, 1975), radiosonde temperature reports were combined with all balloon-borne winds, including pilot-balloon observations of wind only. Thus it is not possible to directly compare the number of radiosonde temperatures with satellite temperatures.

The second column of Table 1 depicts the further expansion of the data data base from 1975 to 1980, reflecting the activities associated with the Global Weather Experiment. The data counts represent averages over 14 analysis times during the period 15-22 February 1980. All categories have shown some increase, and one additional data source has been added - fixed and drifting buoys, largely in the Southern Hemisphere. With respect to remote temperature profiles, the 50% increase is actually an underestimate in that it includes only one of the planned pair of polar-orbiting sounder spacecraft. The 867 soundings of Table 1 were produced by the NOAA-6 spacecraft launched June 27, 1979. The other member of the pair, TIROS-N

Table 1. Summary by data type of reports received at NMC within 10 hours after 0000 GMT or 1200 GMT. Values are averages over 14 analysis times each from November 1975 and February 1980. In the original source for 1975 (McDonell, 1975) rawindsonde reports were not divided into temperature and wind reports.

Data Type	1975	1980	% Change
Surface (land)	4120	4616[1]	+12
Surface (ship)	673	740[1]	+10
Surface (buoys)	0	251	--
Radiosonde Temp.	--	678	--
Satellite Temp.	548	867[2]	+58
Radiosonde Winds	878	885	+ 0.5
Satellite Winds	830	875[3]	+ 5
Aircraft Winds	592	753	+27

1 Does not include data rejected by preliminary quality control; increase is due to expanded processing capabilities.

2 Includes data from only NOAA-6; TIROS-N inoperative during this period.

3 Does not include data from European Space Agency's Meteosat stationed at 0°.

was launched in June 1978 but was malfunctioning during the period covered by the second column of Table 1. With both satellites functioning, the appropriate entry in Table 1 would have been approximately double that actually shown. Similarly, the number of cloud-motion wind vectors is somewhat diminished by the failure of Meteosat, the geostationary satellite operated by the European Space Agency. Prior to December 1979, NMC typically received 300-400 additional reports form this source. Finally, note the 27% increase in aircraft reports in 1980 as compared to 1975. Part of this increase reflects about 100 reports per analysis time from wide-bodied commercial jet aircraft equipped with Aircraft to Satellite Data Relay (ASDAR) devices. These transmit meteorological information measured by the aircraft's sensors and navigation system automatically to ground stations via relay by geostationary satellite. The data thus obtained is highly accurate, reliable, and very timely. Improved processing capabilities account for the remainder of the increase in receipt of aircraft reports.

Table 1 clearly indicates that the global data base is definitely no longer homogeneous. While the radiosonde network remains the cornerstone in the Northern Hemisphere, the other elements are increasingly important. In the Southern Hemisphere, radiosonde data represent only a minor fraction of the data base. Remotely-sensed data and aircraft reports, together with the buoy program, are the dominant elements. Their presence permitted NMC to extend operational analysis procedures to the entire globe in September 1974, and to begin regular global numerical predictions to several days in October 1977.

III. Main Components of the Data Base in 1980

Surface Data

Land Stations: Figure 1 depicts the typical coverage of surface synoptic reports that arrive at NMC prior to 10 hours after observation time. The average number of such reports available to the NMC assimilation system is about 4600. As noted in Table 1, however, this number reflects those reports accepted after the first level of quality control; a greater number actually arrive at NMC. Receipt of these reports varies but little with time of observation; i.e., there are typically as many reports at 1800 GMT as at 1200 GMT.

The distribution is generally adequate in the Northern Hemisphere; perhaps more than adequate in densely-populated areas such as western Europe. However, the distribution is much less dense in the tropics and Southern Hemisphere (Figure 2).

For purposes of numerical weather prediction, the surface pressure is the observed parameter of greatest utility; temperature and wind are too susceptible to non-representative local effects. The NMC assimilation system uses observed station pressure in preference to pressure reduced to sea level, although reduced pressure will be used if the station's elevation is less than 500 m and station pressure is not reported. Observational errors are assumed to be random in character.

Ships: Figures 1 and 2 also depict typical coverage of surface ship reports over the world's oceans. The density of ship data is much less than that of land stations. Of the average 740 ship reports indicated in Table 1, about 600 are from the Northern Hemisphere. The distribution of ship reports with observation time varies markedly in contrast to the land reports. Table 2 illustrates the variation of the number of ship reports with time of observation. This variability is distributed spatially: for example, the maximum number of reports in the North Atlantic usually occurs at 1200 GMT, whereas in the Pacific it occurs at 0000 GMT.

Table 2. Typical variation in the number of ship observations as a function of the time of observation, 2 March 1980.

	0000 GMT	0600 GMT	1200 GMT	1800 GMT
NUMBER	701	527	695	617

Some of this variation is due a longer data cutoff time at 0000 GMT and 1200 GMT - about 9h - as opposed to a 4h cutoff at 0600 GMT and 1800 GMT.

Most of the land weather stations which make and transmit surface-based observations are operated by governments. In contrast, most of the ships are operated by commercial interests although equipment may be furnished by government. The primary purpose of the commercial ship is not to serve as a meteorological observatory but rather to transport

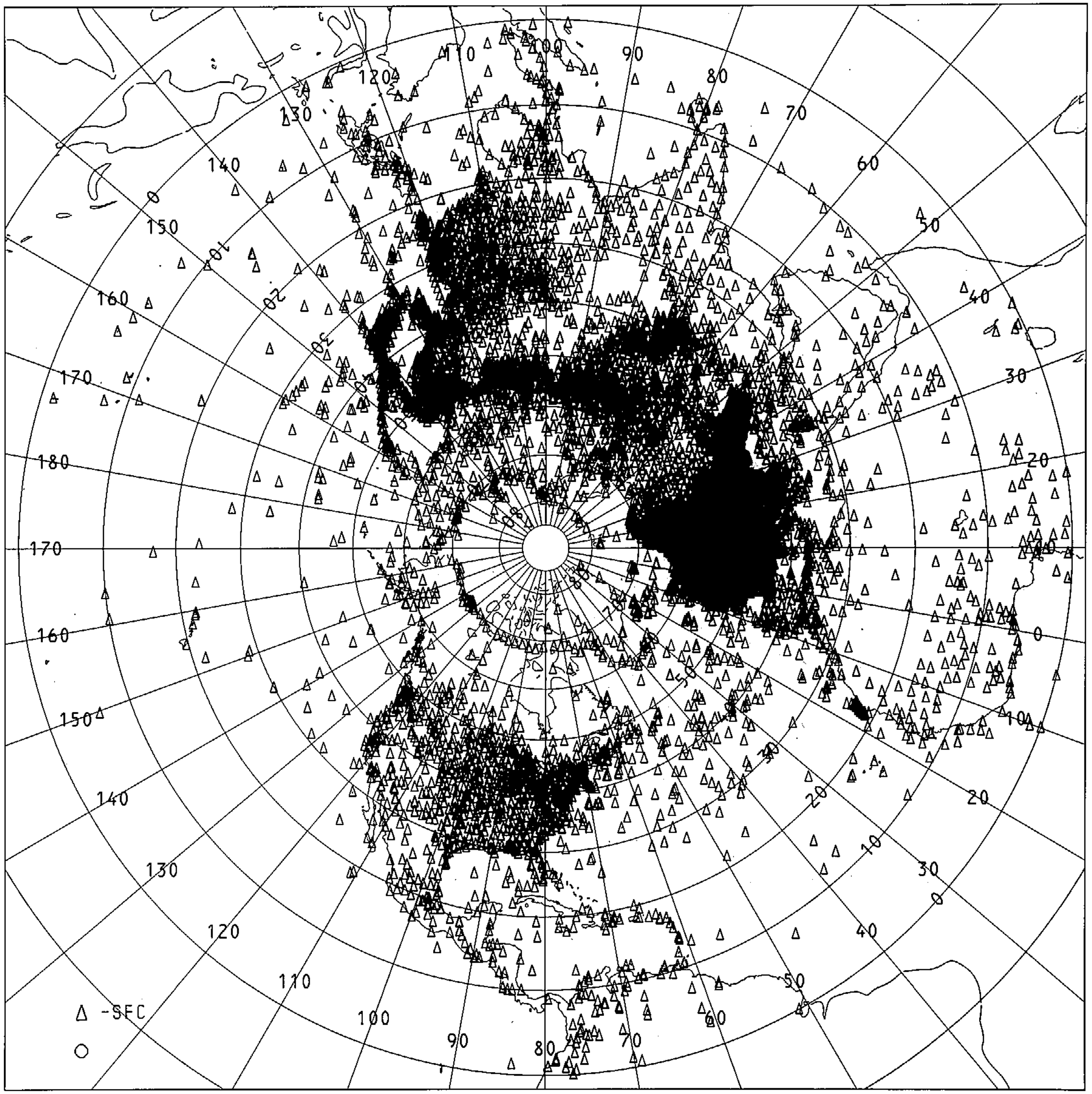


Figure 1. Northern Hemisphere surface data coverage 1200 GMT 21 October 1979.

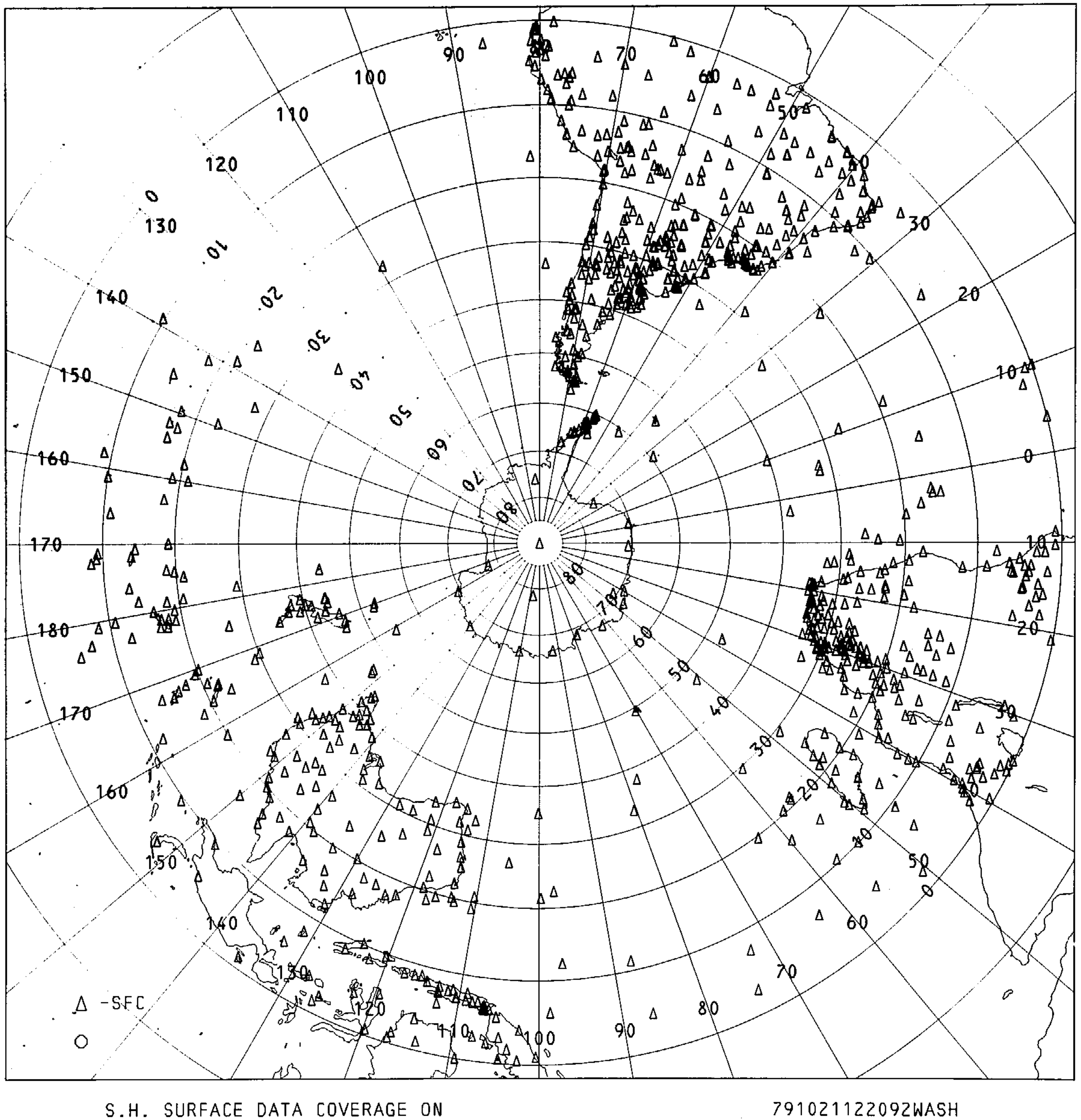


Figure 2. Southern Hemisphere surface data coverage 1200 GMT 21 October 1979.

passengers or freight. Consequently, weather observations have a lower priority with respect to claims on the time of the crew. One result of this is the variation in the number of observations depending on the time of day; e.g., 1200 GMT is local midnight in mid-Pacific when most crew members prefer to be sleeping. Another consequence is the somewhat lower reliability of ship observations.

Buoys: The United States presently maintains and operates 21 fixed (anchored) buoys in the coastal and offshore waters under its jurisdiction. Wind direction and speed, and pressure measurements are transmitted by geostationary satellite to NMC for processing and entry in the data base. These have proven quite useful for enhancing the analysis of near-surface conditions in coastal areas.

For large scale numerical weather prediction, the impact of the drifting buoys in the Southern Hemisphere has been very significant, owing largely to the extremely poor coverage by ships. Figure 3, taken from Fleming, et al., (1979), shows the location of the drifting buoys on one day during the first Special Observing Period of the GWE. It will be noted that the density of reports still does not bring coverage up to the same level as in the Northern Hemisphere, but it nevertheless represents a substantial improvement. The drifting buoys provide pressure and sea-surface temperature measurement, relayed via orbiting satellite to ground stations for processing and transmission on communications circuits.

Sea Surface Temperature: Ships also provide some information on the sea surface temperature, the analysis of which is required as a boundary condition in numerical weather prediction. Additional information is provided by temperature vs. depth (BATHY) and temperature/salinity vs. depth (TESAC) soundings. To supplement these *in situ* measurements, radiometric data from orbiting satellites are also used. Gemmill and Larson (1979) indicate that in any 24-hour period, reports from 2800 ships, 80 fixed buoys, 200 drifting buoys, and 150 BATHY/TESAC instruments can be anticipated, along with some 40,000 satellite estimates of sea surface temperature. The data from these sources differ considerably in their characteristics. Ships, for example, measure water temperature at intake ports several meters below the vessel's water line. Gemmill and Larson estimate the error of the observations as 1.5°C. Buoy measurements are within one meter of the surface, and are estimated to be accurate to within ± 0.2°C. These data, along with the BATHY/TESAC reports, are the most accurate of the several sources. Unfortunately, the coverage they afford is limited. On the other hand, coverage is the primary advantage of satellite sea surface temperature estimates. Gemmill and Larson assign the same error level to satellite data as to ships, but note that the former often exhibit systematic errors over fairly large areas. This renders difficult blending the *in situ* and remote measurements into a coherent representation of the sea surface temperature field.

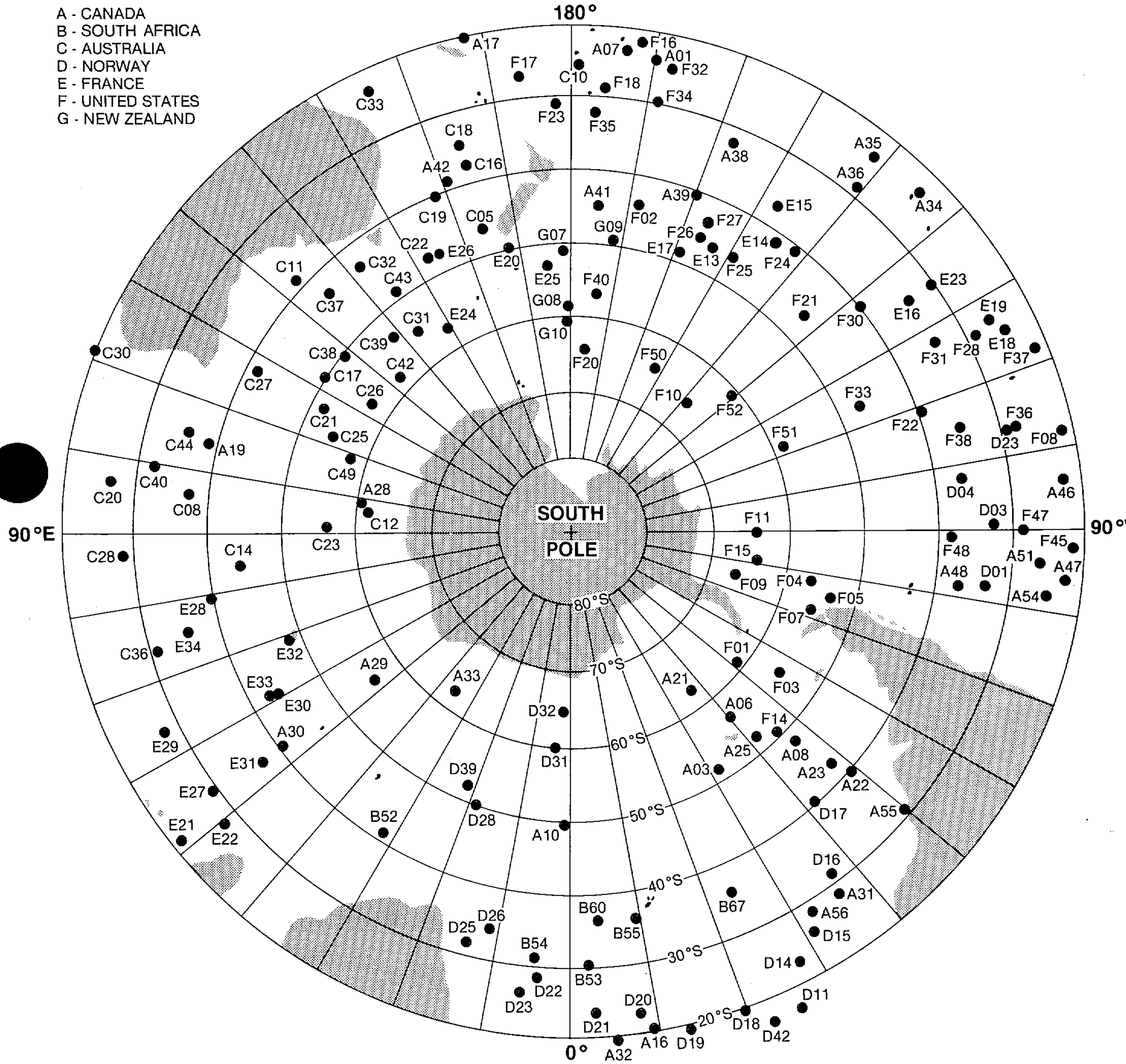


Figure 3. Location of operational drifting buoys in the Southern Hemisphere, 15 February 1979. After Fleming et al., (1979).

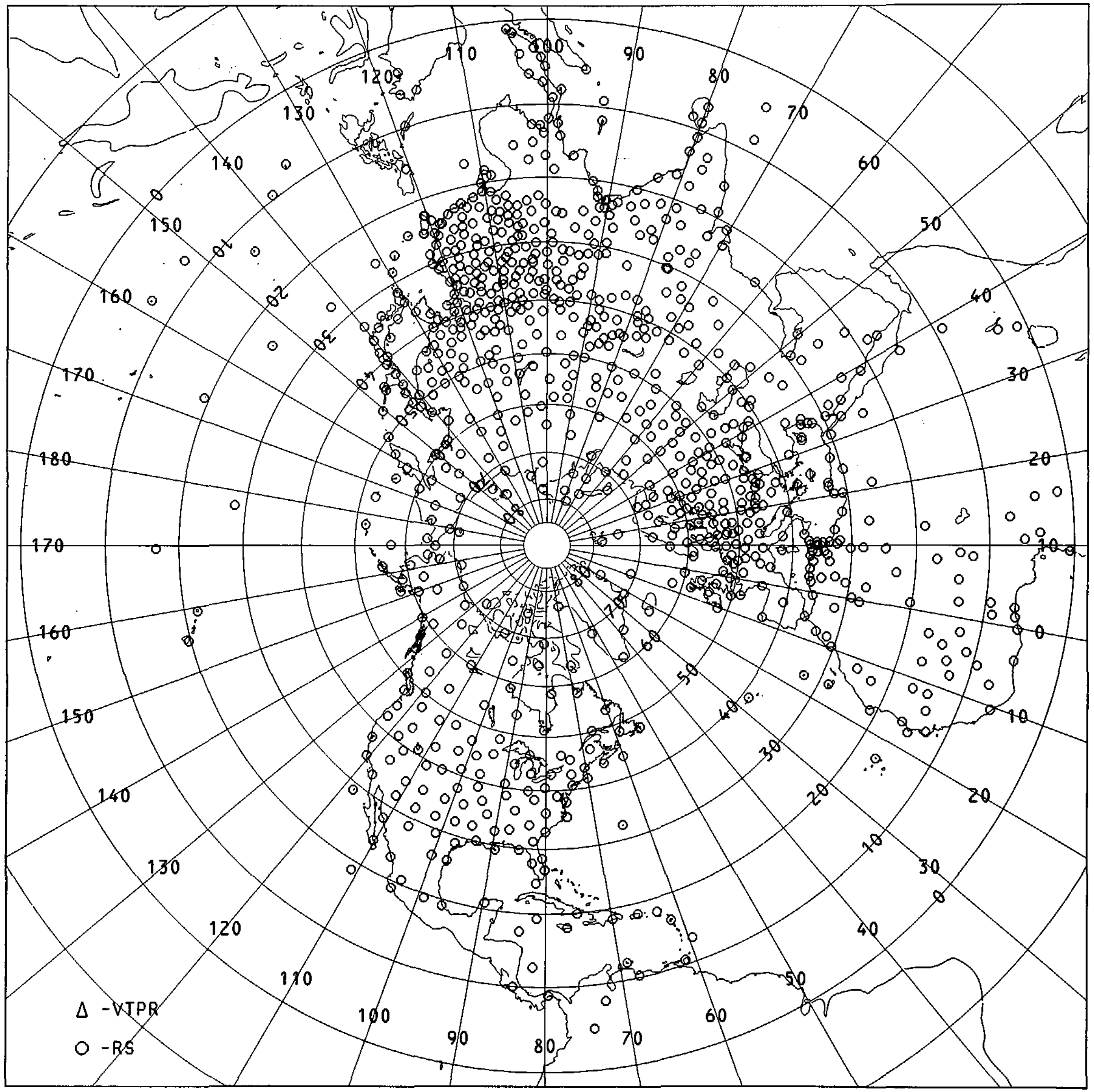
Upper Air

Radiosondes: Figures 4 and 5 depict the typical coverage by radiosonde observations in the Northern and Southern Hemispheres, respectively. Most radiosondes are launched near 0000 GMT and 1200 GMT; a few stations make four launches daily. Note that the coverage is most dense over Europe, and almost nonexistent over the oceans, where only a few island stations and fewer ships provide radiosonde information. The ships are not depicted in the diagrams; they are extremely expensive to maintain, and are gradually being phased out.

The radiosonde report gives temperature and wind as functions of pressure. Since the balloon ascends at a relatively slow rate (approximately one hour to reach 100 mb), the measured profiles of temperature and wind contain more detailed information than can profitably be used by current large-scale prediction models. That is, the radiosonde senses some scales of phenomena which are too small to be represented by the resolution of the prediction model. Small-scale, high frequency variations in the data are treated as errors due to sampling, and are combined with errors in the measurements themselves to give a total observational error. Temperature errors are generally considered to be random; the error at one location is not well-correlated with the error and any other location, nor with the true atmosphere. Wind errors, however, tend to vary with wind speed: higher wind speeds inevitably involve larger wind errors. The quality and reliability of both temperatures and wind reports from at least United States radiosondes has improved considerably in recent years, as a result of radar tracking and the use of on-site mini-computers.

One recent study (Bruce et al., 1977) concluded that the combined temperature observational error for radiosondes released over the White Sands Missile Range (about 200 km^2) is near 0.8°C with only slight variation in the vertical to approximately 100 mb. With respect to observed winds, a study by Bauer (1976) has determined that the mean absolute error in a sample from three observation times in October 1974 averaged approximately 5 $msec^{-1}$. It is unfortunate that Bauer's results are expressed in terms of mean absolute error, for there is no direct way of converting these numbers to observational error variances. For purposes of correcting colocation statistics - differences between approximately colocated radiosonde winds and cloud-motion vectors on aircraft reports - RMS errors of 1.8 $msec^{-1}$ and 5.9 $msec^{-1}$ per component are assigned to low--and high-level radio winds, respectively.

Remote Soundings: Figures 6 and 7 show the coverage afforded by a single polar-orbiting observation platform equipped with a sounding radiometer. The coverage shown is from both TIROS-N and NOAA-6 satellites, and includes soundings made within $\pm$ 3 hours of 1200 GMT 21 October 1979. When both the TIROS-N and NOAA-6 satellites are in operation, approximately global coverage is available each 12 hours. In low latitudes, there are some gaps between successive swaths of data, and considerable overlap occurs in high latitudes. The latter necessitates some reduction or "thinning" of the data prior to ingestion in the NMC assimilation system. Furthermore, remote sounding data over continental areas are not presently used in the assimilation system. The average



N.H. RADIOSONDE DATA COVERAGE ON 791021122092WASH

Figure 4. Northern Hemisphere radiosonde coverage, 1200 GMT 21 October 1979. Ocean Station Vessels are not included.

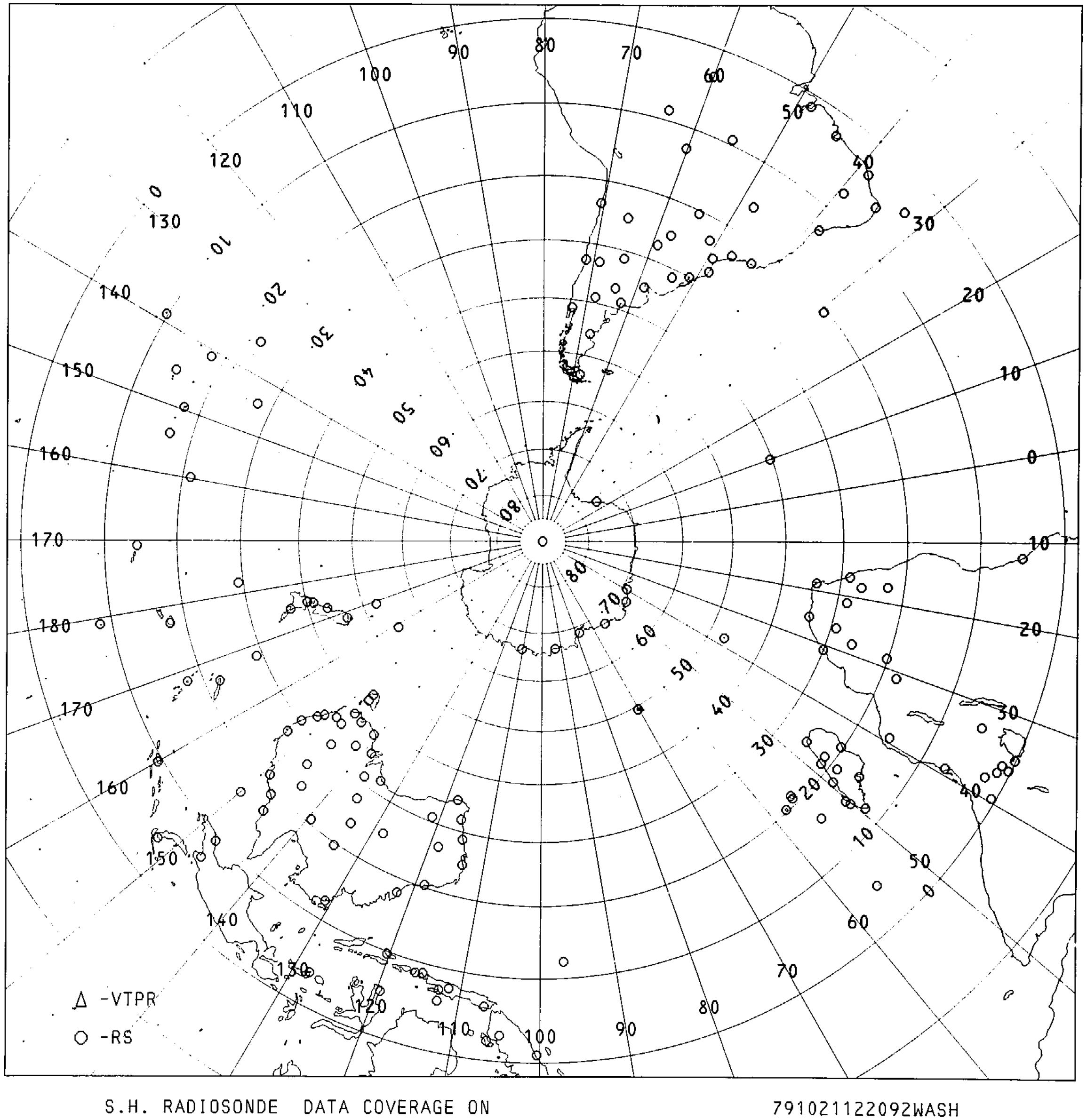


Figure 5. Southern Hemisphere radiosonde coverage, 1200 GMT 21 October 1979.

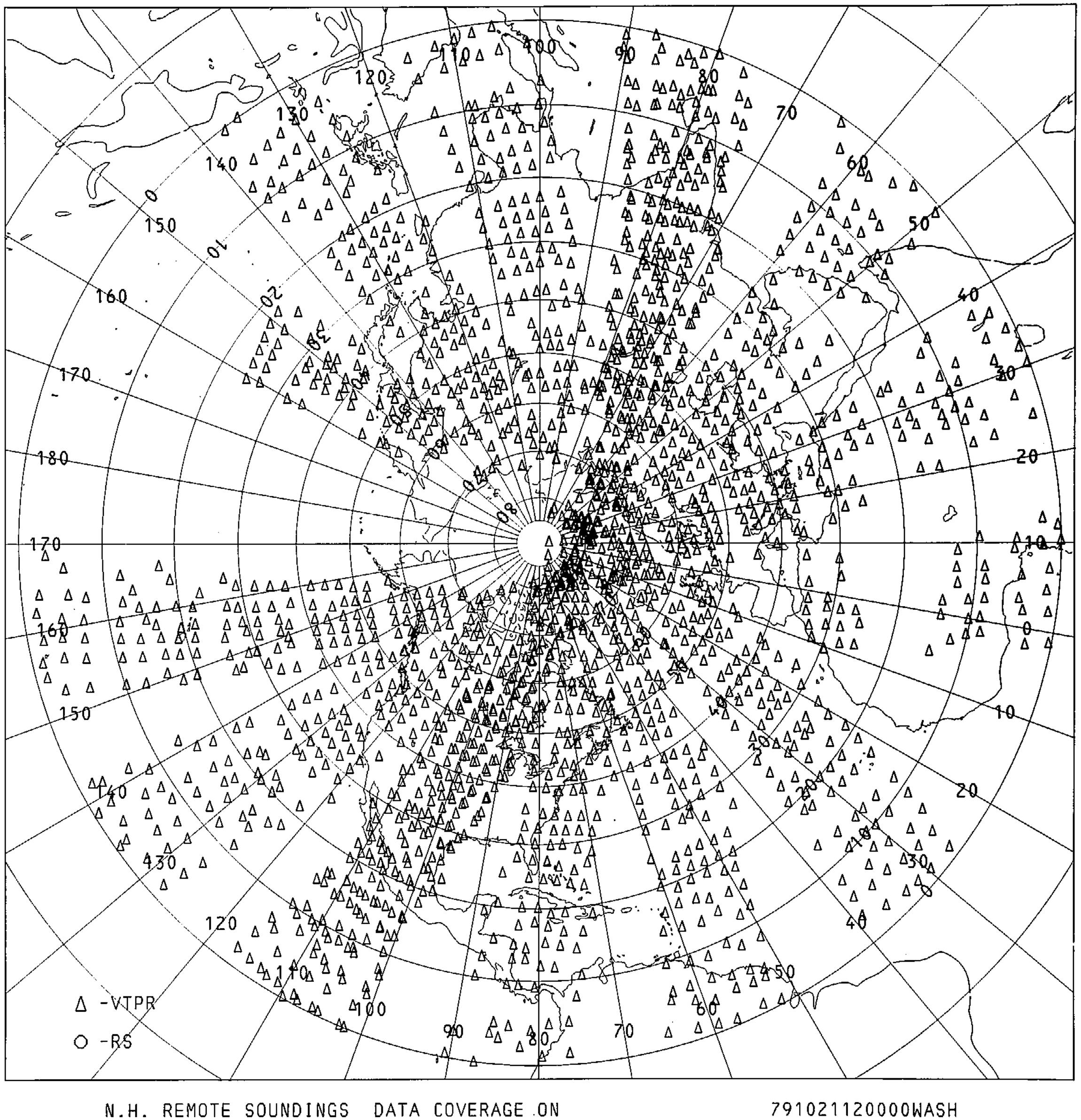
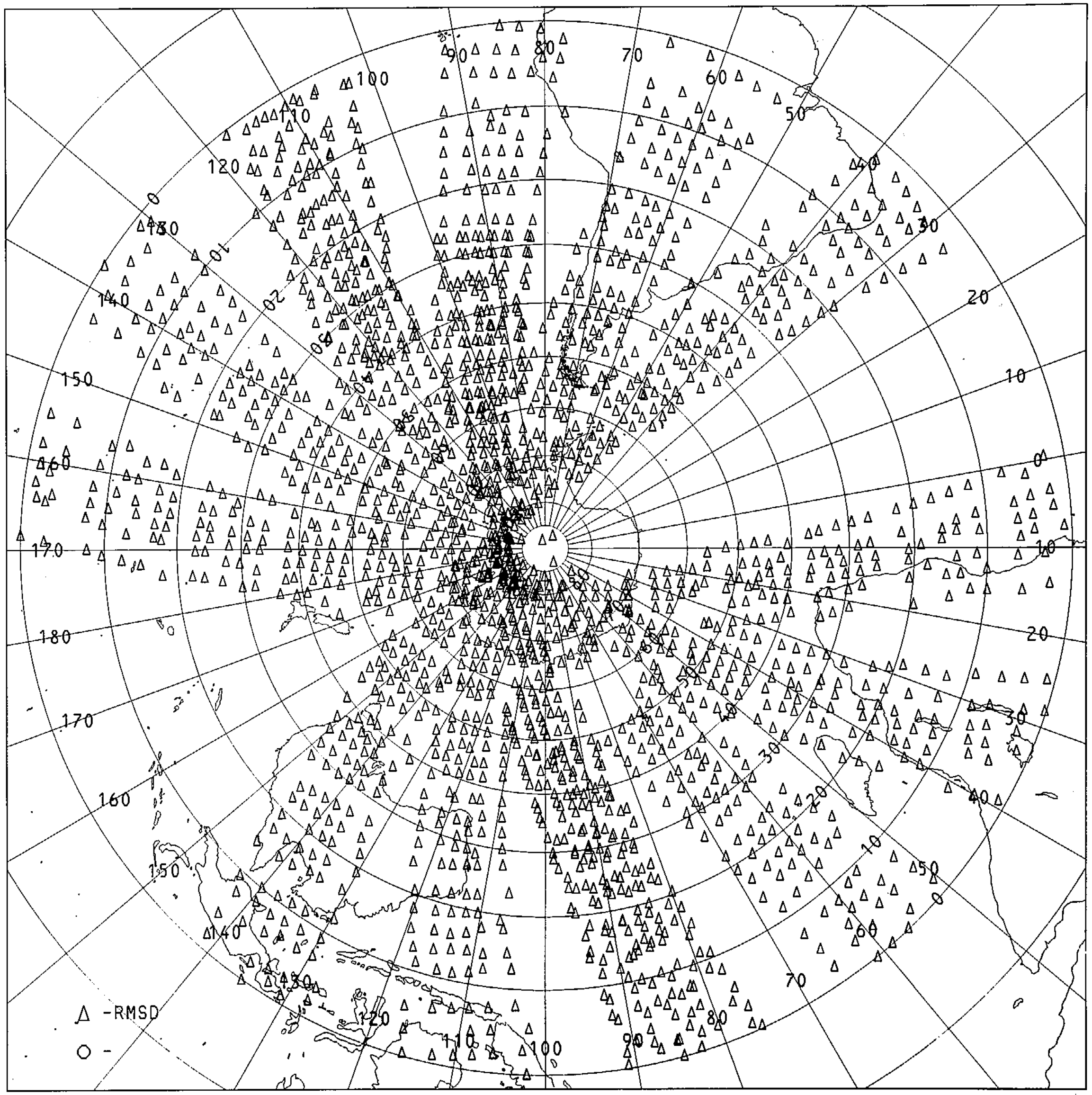


Figure 6. Northern Hemisphere remote sounding coverage from TIROS-N and NOAA-6 during 0900 GMT - 1500 GMT 21 October 1979.



S.H. REMOTE SOUNDINGS DATA COVERAGE ON 791021120000WASH

Figure 7. Same as Figure 6, but for the Southern Hemisphere.

number of remote soundings - 867 - that appears in Table 1 reflects an underestimate of the number of soundings actually available, even from one satellite. Ground processing of the data generally is completed by 1.5 to 3 hours after the actual time of observation. Variability in this lag is mostly due to communications delays in transmitting the data from the satellite to a ground readout station and then to the Central Computing Facility in Suitland, Maryland for actual processing.

The retrieval process - transforming the measured radiances at several frequencies into estimates of layer-mean temperatures - is statistical in character. Although it is possible to use a physical method - numerical inversion of the radiative transfer equation (Wark and Fleming, 1966) - the statistical approach is presently thought superior. Basically, the retrieval process relates the mean virtual temperature for any layer to a linear combination of the observed radiances I_k:

$$T_\ell = \sum_{k=1}^{K} a_k I_k$$

The coefficients a_k are determined from a dependent set of data consisting of approximately colocated radiosondes and remote soundings - within 3 hours in time and 1° latitude in space. The colocation base is stratified by latitude and is updated weekly; it therefore reflects seasonal variability. Coefficients determined from this dependent sample are then applied to independent radiance measurements to obtain temperature profiles.

In practice, it is not so simple. The effects of clouds must be accounted for as must noise in the radiance measurements. The orbiting infrared radiometer senses radiation from a relatively small area - a circle about 30 km in diameter (Smith, et al., 1979) - as it scans from side to side along the orbital path. Several such measurements are combined in each sounding, as one method of reducing random noise in the measurements. In addition, the retrieval method currently in use actually performs a least-squares fit of the data to the eigenvectors of the temperature/radiance covariance matrix. This also tends to reduce noise.

Cloudiness exerts a powerful influence on the efficacy of indirect sounding techniques. Infrared radiation is profoundly affected by the presence of clouds. In early remote sensing satellites, this effectively restricted the retrievals to clear areas. Since clear areas tend to be associated with anticylones and quiescent weather, remote soundings were obtainable mostly where they were least needed. The addition of several microwave frequencies on more recent sounders has alleviated the problem to some extent, since microwave radiation is not affected by non-precipitating clouds.

For the present TIROS-N/NOAA-6 sounding satellites, retrievals are made in three different ways depending upon the amount of cloudiness present, which is sensed internally from the radiance measurements themselves. For clear skies, only the infrared frequencies are used, and the retrieval proceeds as outlined in previous paragraphs. When a certain amount of cloudiness can be detected, a correction to the cloud-contaminated measurements is made, based on nearby clear measurements. If cloudiness is sufficiently extensive that no clear measurements can be found close at hand, then information from the microwave frequencies is used. These "clear", "partly cloudy", and "cloudy" retrievals may therefore be expected to have rather different error characteristics. As a general rule, the clear retrievals are most accurate since they are performed under optimum conditions. Clear retrievals constitute as much as 90% of the total over oceans in Northern Hemisphere summer.

The second-path retrievals constitute a much smaller percentage of the total. Although it might be expected that "partly cloudy" retrievals would be at least slightly less accurate than clear ones, this has not been conclusively demonstrated. In part for this reason, the error statistics presented in Table 3 do not reflect second path retrievals as a separate category.

Cloudy retrievals can total as much as 65% of all retrievals in Northern Hemisphere winter (Phillips, 1980). This is unfortunate, since these are less accurate, and exhibit less resolution. Nevertheless, cloudy retrievals do provide some information where none was available before.

Figure 8 shows a typical remote temperature sounding compared with a nearby radiosonde profile, after Phillips, et al. (1979). Note that the details of the temperature structure - shallow inversions, etc. - are not resolved by the remote sounding. This reduced vertical resolution is further illustrated by the vertical cross-sections of Figure 9, after Tracton, et al. (1980). The upper diagram, produced from radiosonde data alone, clearly depicts strong baroclinic zones which are greatly smoothed in the lower diagram.

Coarse vertical resolution, cloud contamination, and the innate conservative nature of the statistical retrieval method are major contributors to observational errors in remote soundings. Table 3, taken from Phillips et al., (1979) gives a statistical estimate of the errors based on colocations with radiosondes. It will be noted that the clear retrievals have fairly large errors in the lowest layers and around the tropopause, mostly reflecting vertical resolution. In the middle troposphere, however, colocation differences between remote soundings and radiosondes are not much larger than differences between radiosondes over the colocation "window". The cloudy retrievals, on the other hand, show very large differences with radiosondes especially in the lowest layers. Phillips (1980) has recently suggested that a combination of rain contamination of the retrievals and a slant toward continentality of the colocation base used to determine the coefficients may account for the large low-level errors, especially the large mean error.

The mean error illustrates another characteristic of remote soundings; the data tend to display local "biases", or non-zero mean errors, over appreciable geographic areas. Statistically, this appears as a spatial correlation of the remote sounding errors. If a remote sounding temperature for a given layer can be identified as too warm, for example, then there is a high probability that its adjacent neighbors will also be too warm. Schlatter (1980) has investigated the errors of TIROS-N soundings and found a definite spatial correlation of errors which decays exponentially

Table 3. Differences between TIROS-N retrievals and colocated radiosondes layer-mean temperatures for seven marine station in the Northern Hemisphere extratropics: Lajes (39N), 5 ships (47N, 50N, 53N, 57N, 66N), Adak (52N). Averaged over 29 March - 27 April 1979. Extracted from Phillips, et al., (1979).

	Mostly clear		Mostly cloudy	
Pressure layer (cb)	Mean	RMS	Mean	RMS
100-85	-0.6	2.3	1.6	3.1
85-70	-0.2	1.4	1.5	2.4
70-50	0.1	1.7	0.3	2.4
50-40	0.3	1.8	0.3	2.5
40-30	0.8	2.2	0.3	2.4
30-20	0.6	2.0	0.9	1.9
20-10	0.2	1.8	0.0	1.6

from near unity at small separations to zero at about 800-1000 km. Although this detracts from the value of the data when performing an analysis of the temperature, Bergman (1978), Seaman (1977) and others have shown that spatially correlated temperature errors are more helpful to an analysis of the motion field, since gradient information is used.

Cloud-Motion Wind Vectors. Even so, observations of the mass field are no substitute for wind observations in the motion analysis. This is especially true in low latitudes where the mass-motion balance is much more complicated than geostrophic or balance-equation constraints suggest; but it is also true at mid- and high-latitudes where the ageostrophic portion of the wind field plays an important role in accurate forecasting. Consequently, the cloud-motion wind vectors from geostationary satellite cloud imagery have proven a valuable source of data.

Figures 10 and 11 show the coverage afforded by cloud-motion wind estimates, all levels combined for the Northern and Southern Hemisphere, respectively. Note that coverage extends only up to about 50° latitude. Note also that major gaps exist. These are in part due to a lack of cloud targets in some areas. However, the most noticeable gap extends from about the Greenwich meridian eastward for roughly 100 degrees. This is a result of the failure of the European Space Agency's Meteosat, which was positioned at 0° longitude, and the not-yet-operational status of a geostationary satellite at 58°E. Figure 12 depicts the coverage expected when all five satellites are functioning.

The most noticeable characteristic of this source of data is its bimodal distribution in the vertical. For the most part, low-level vectors are obtained by tracking cumulus clusters which reflect the wind field between 700 mb and 800 mb. High-level vectors are exclusively determined from cirrus elements. Experience has shown that relatively little mid-level cloudiness occurs unobscured by higher cirrus. Hence, the data tend to be available only around 850 mb and 250 mb.

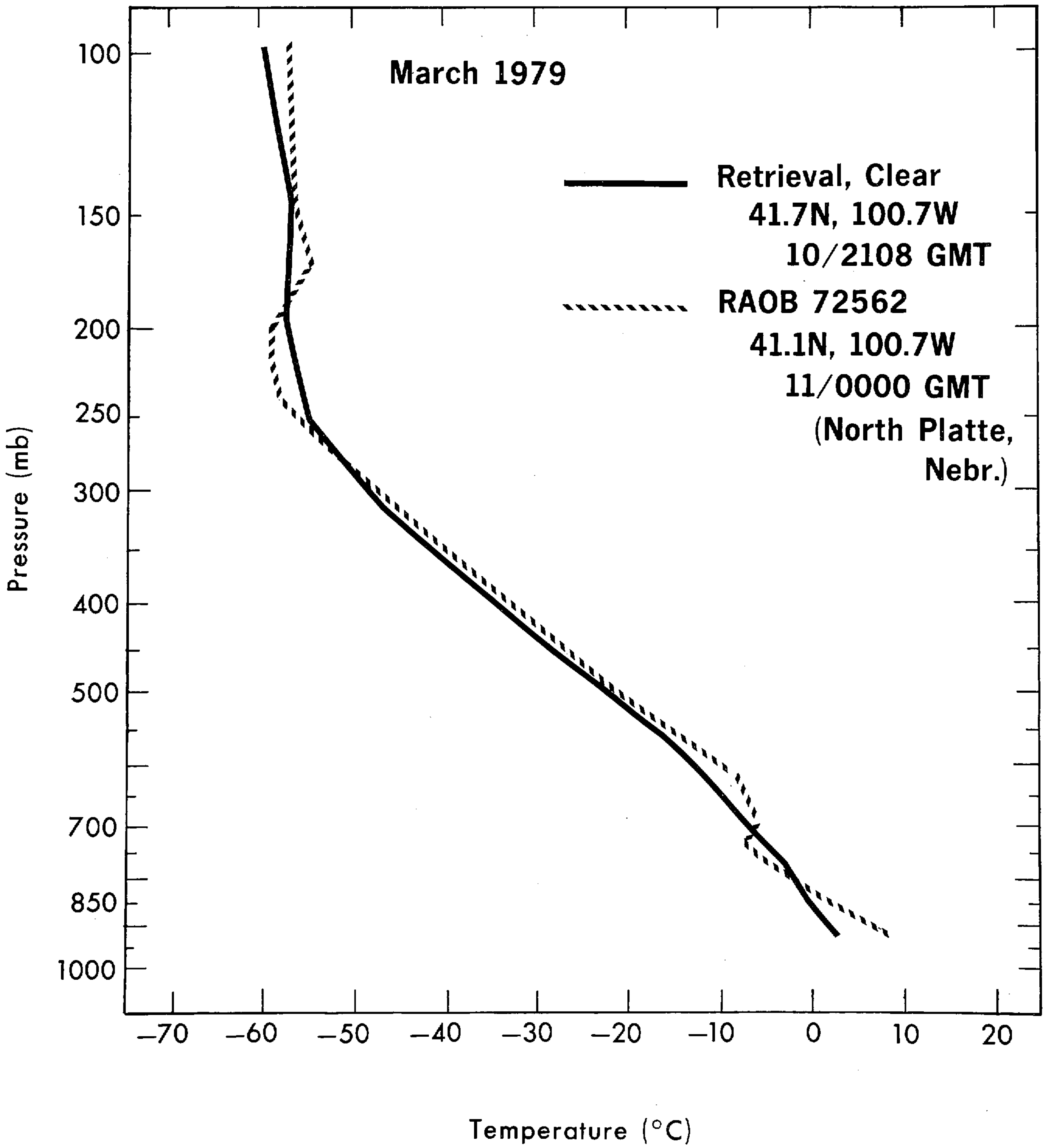
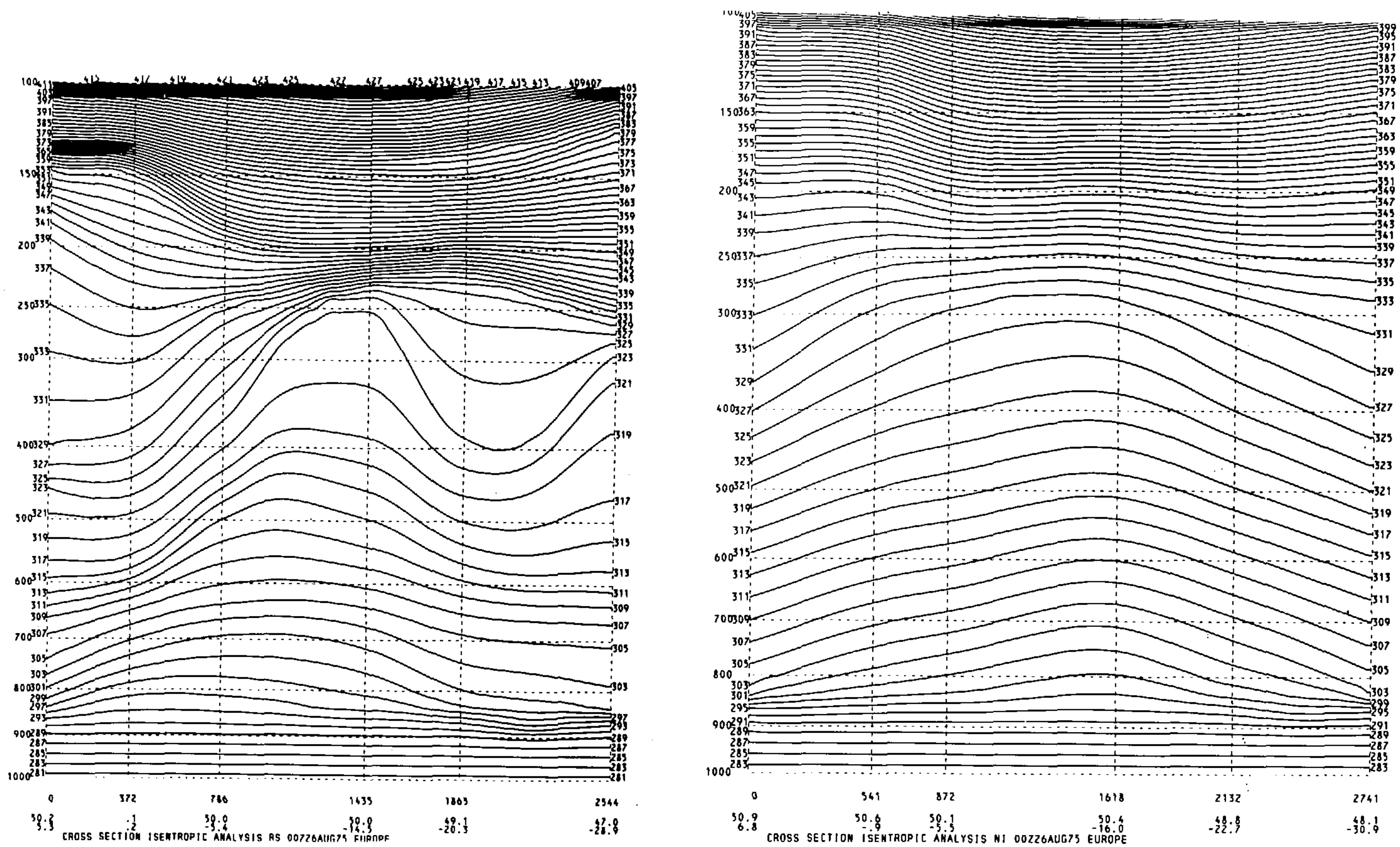


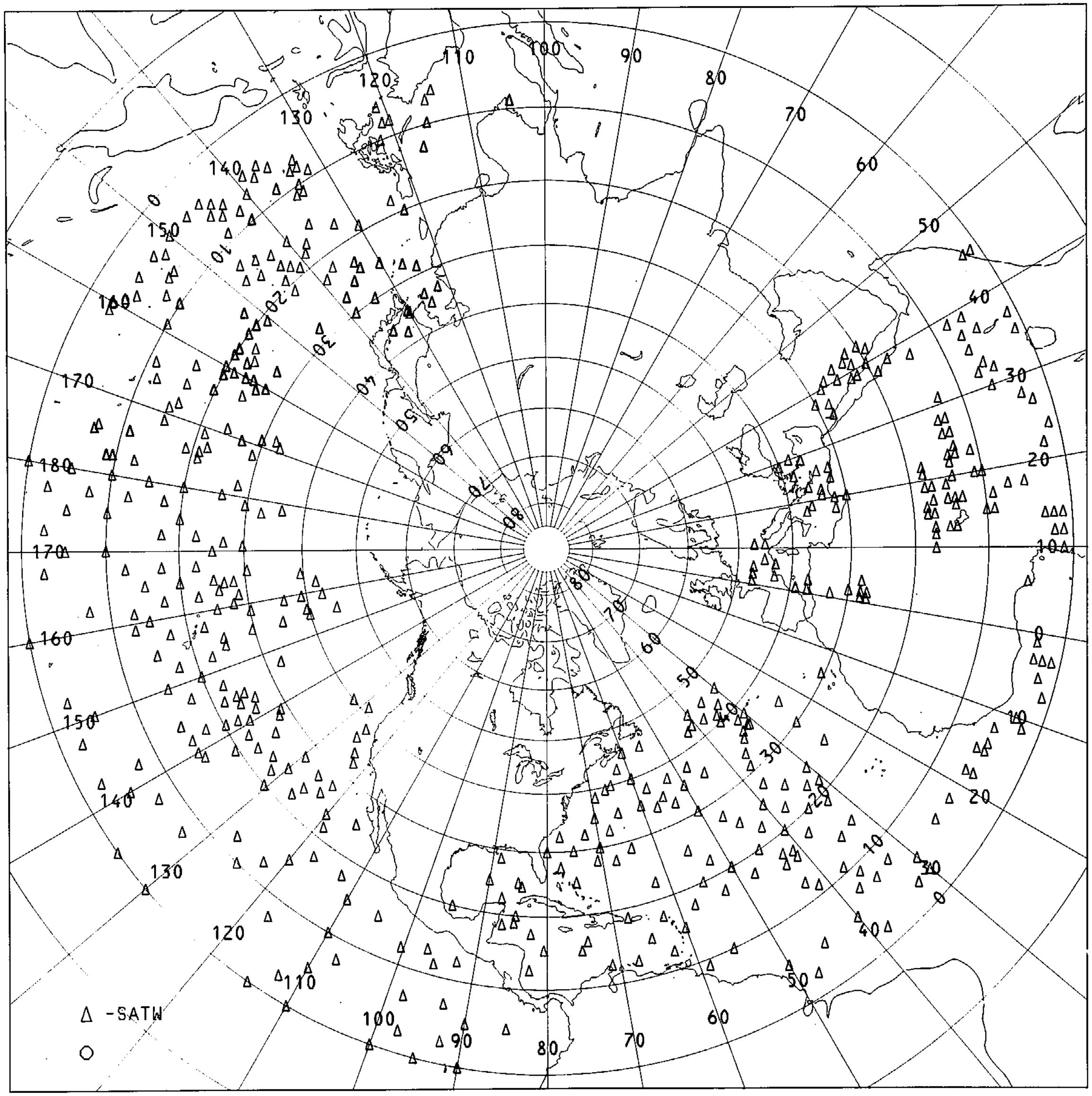
Figure 8. Comparison of colocated remote sounding and radiosonde temperature profiles, after Phillips et al., (1980).



a

b

Figure 9. Isentropic cross-section analysis for 0000 GMT 26 August 1975: a. using radiosonde data only; b. using remote sounding data only. After Tracton et al., (1980).



N.H. SATWIND DATA COVERAGE ON

79102112209ZWASH

Figure 10. Northern Hemisphere cloud-motion wind vector coverage, all levels combined, for 1200 GMT 21 October 1979.

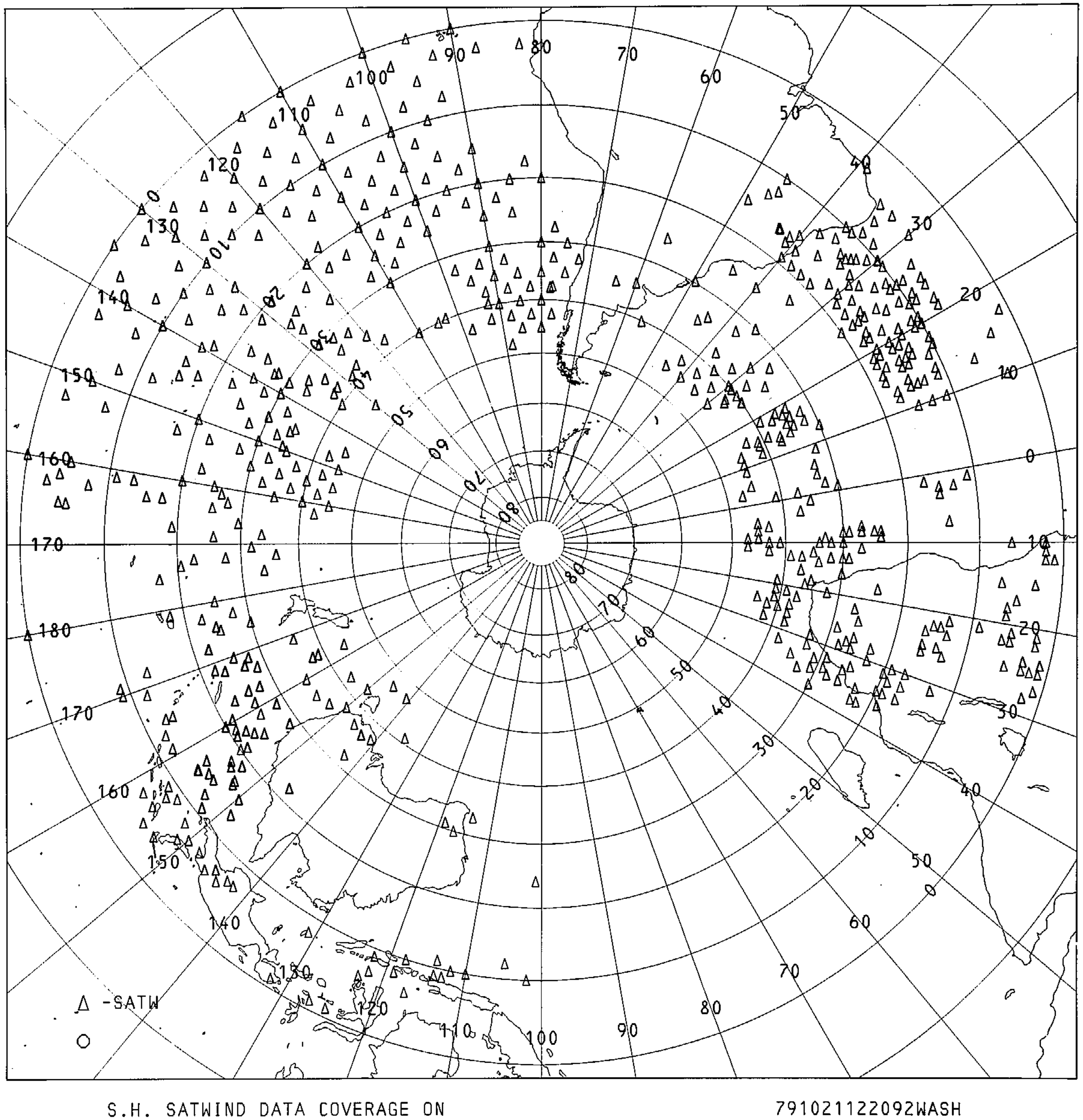


Figure 11. Same as Figure 10, but for Southern Hemisphere.

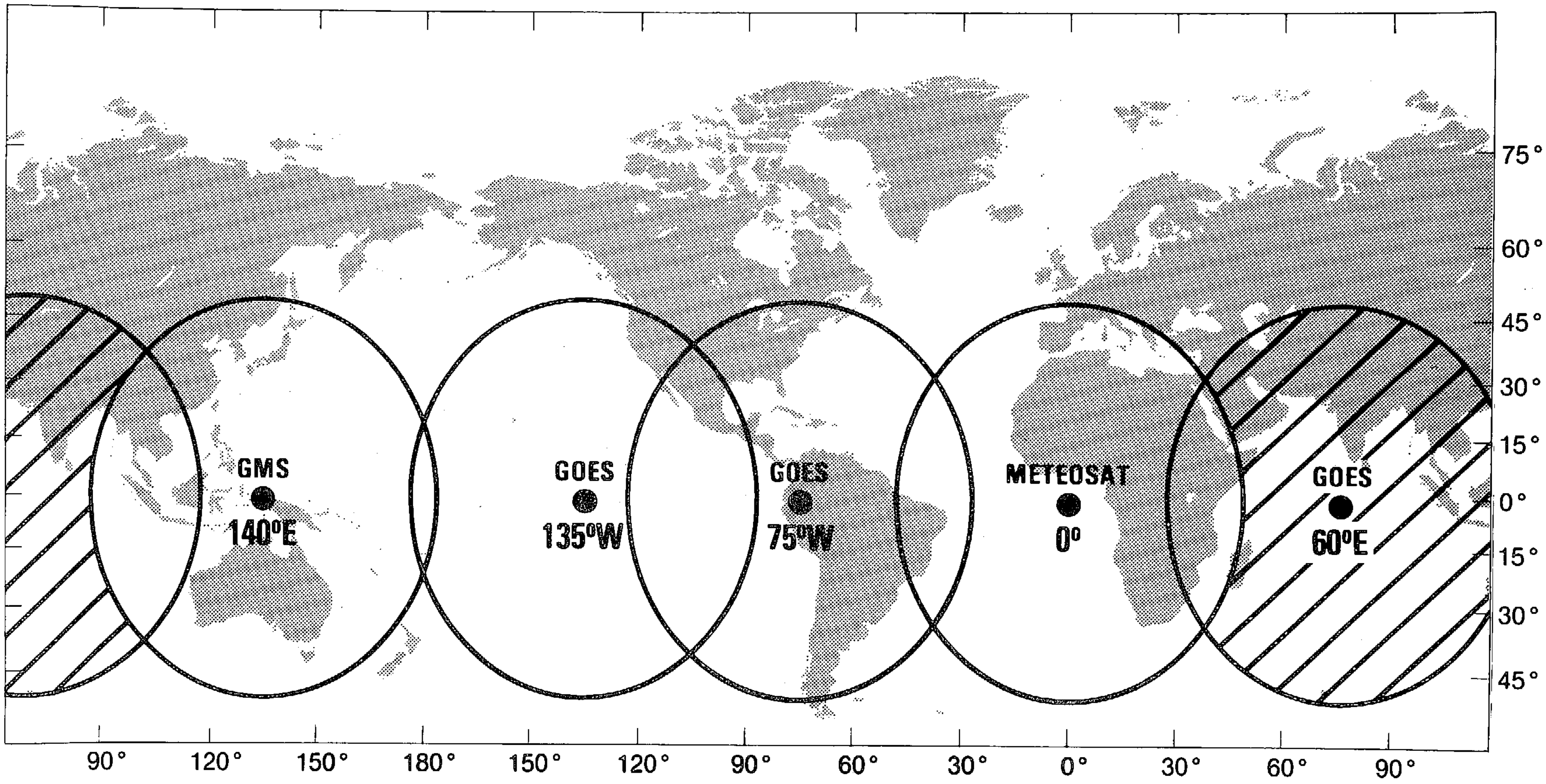


Figure 12. Expected coverage of cloud motion wind vectors when a network of five geostationary satellites is operational. After Fleming et al., (1979).

There are three main sources of error in these data. First are navigational and computational errors. In order to accurately calculate target displacements, the location of the target relative to coordinates on the earth's surface must be known with a high degree of precision. Second, the assumption that the cloud target moves with the prevailing wind introduces some error. Generation or dissipation of targets, or motions which are not reflective of the prevailing wind, adversely affect the wind estimates. In most systems, these errors are minimized by careful target selection and diligent quality control. Third and most important is the altitude assigned to the wind vectors. It is very difficult to decide at what level the apparent motion is occurring to within 100 mb without additional information. Typically, this is supplied by the infrared "brightness temperature" of the cloud target. By matching this temperature with the best available temperature profile, the accuracy of height assignment can be improved. Even so, it remains the chief source of error.

Table 4 provides an estimate of the standard observational errors associated with low- and high-level vectors from four of the satellites, taken from colocations with radiosonde winds over January-March 1979. The RMS differences have been corrected for estimated radiosonde wind errors (1.8 msec^{-1} low level, and 5.9 msec^{-1} high level) by assuming that radio wind errors and cloud wind errors add as squares to form a mean square difference. The values are generally in good agreement with each other except for the error in the Japanese high-level winds. It is believed that the large error is attributable to the altitude assignment technique used in Japan during this period.

These errors are fairly large; but as will be shown in the next lecture, where no other wind data exist cloud-motion wind vectors are quite useful.

Table 4. Estimated observational errors (msec^{-1}) of cloud motion wind components after Morone (1979). Based on colocations with radiosondes over January-March 1979. Colocation window is 3° latitude and 1 hour. Colocation differences have been corrected for assumed radiosonde random errors of 1.8 msec^{-1} (per component) for low level winds (700 mb or lower) and 5.9 msec^{-1} (per component) for high level winds.

Level	European	Japanese	U.S.-West	U.S.-East
low	7.2	6.1	5.4	4.2
high	8.4	13.3	7.0	7.5

Aircraft Data. Along well-established commercial routes, there exists a wealth of wind observations from aircraft. These suffer even more than cloud winds from inadequate distribution in the vertical; most occur in the 100 mb layer between 300 mb and 200 mb. On the other hand, the flight level of the aircraft is known quite accurately. Furthermore, the navigation equipment on modern jet aircraft permits the determination of the wind quite accurately.

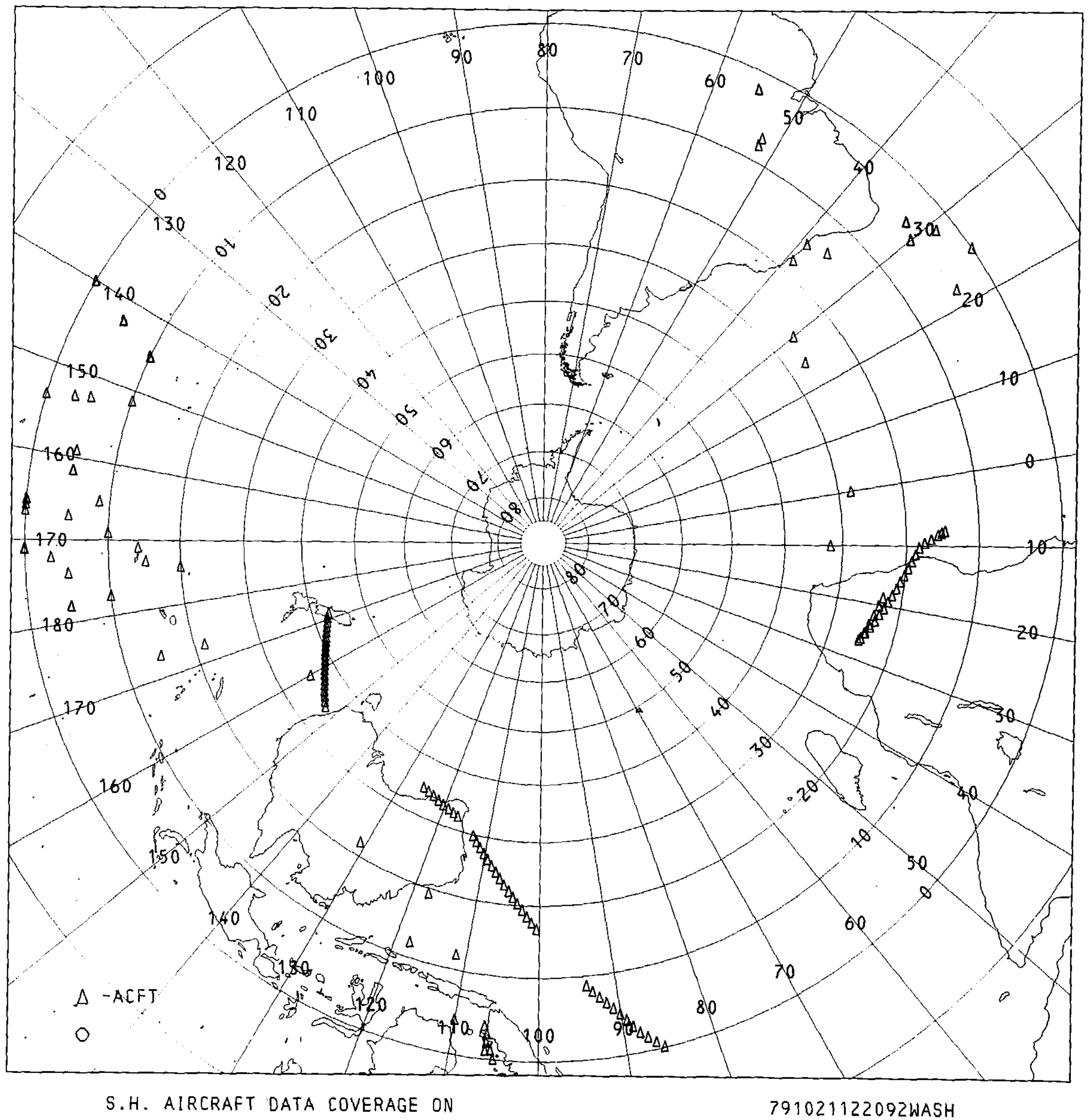


Figure 14. Same as Figure 13, but for Southern Hemisphere.

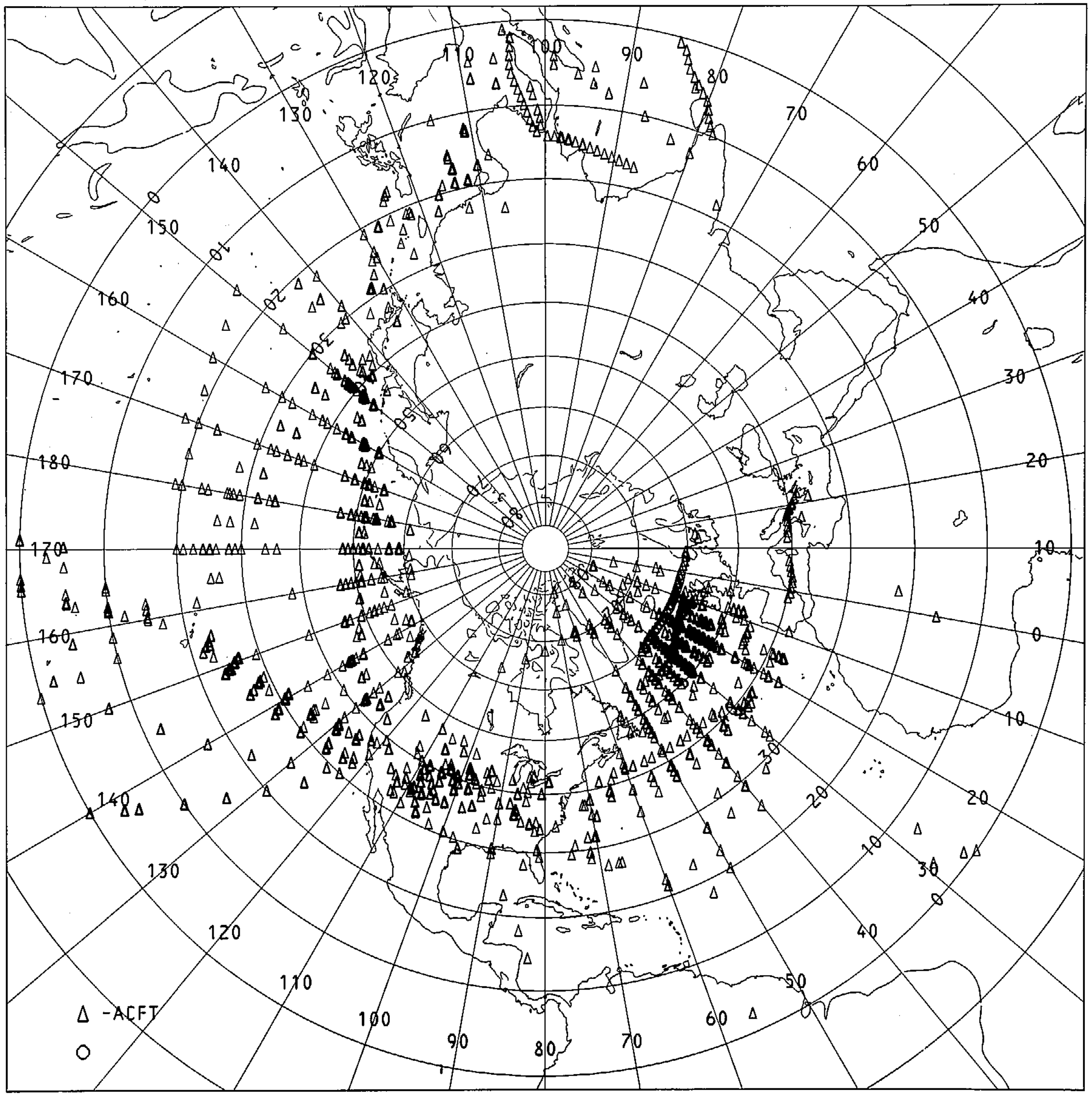


Figure 13. Northern Hemisphere aircraft data coverage during the period 0900 GMT - 1500 GMT 21 October 1979.

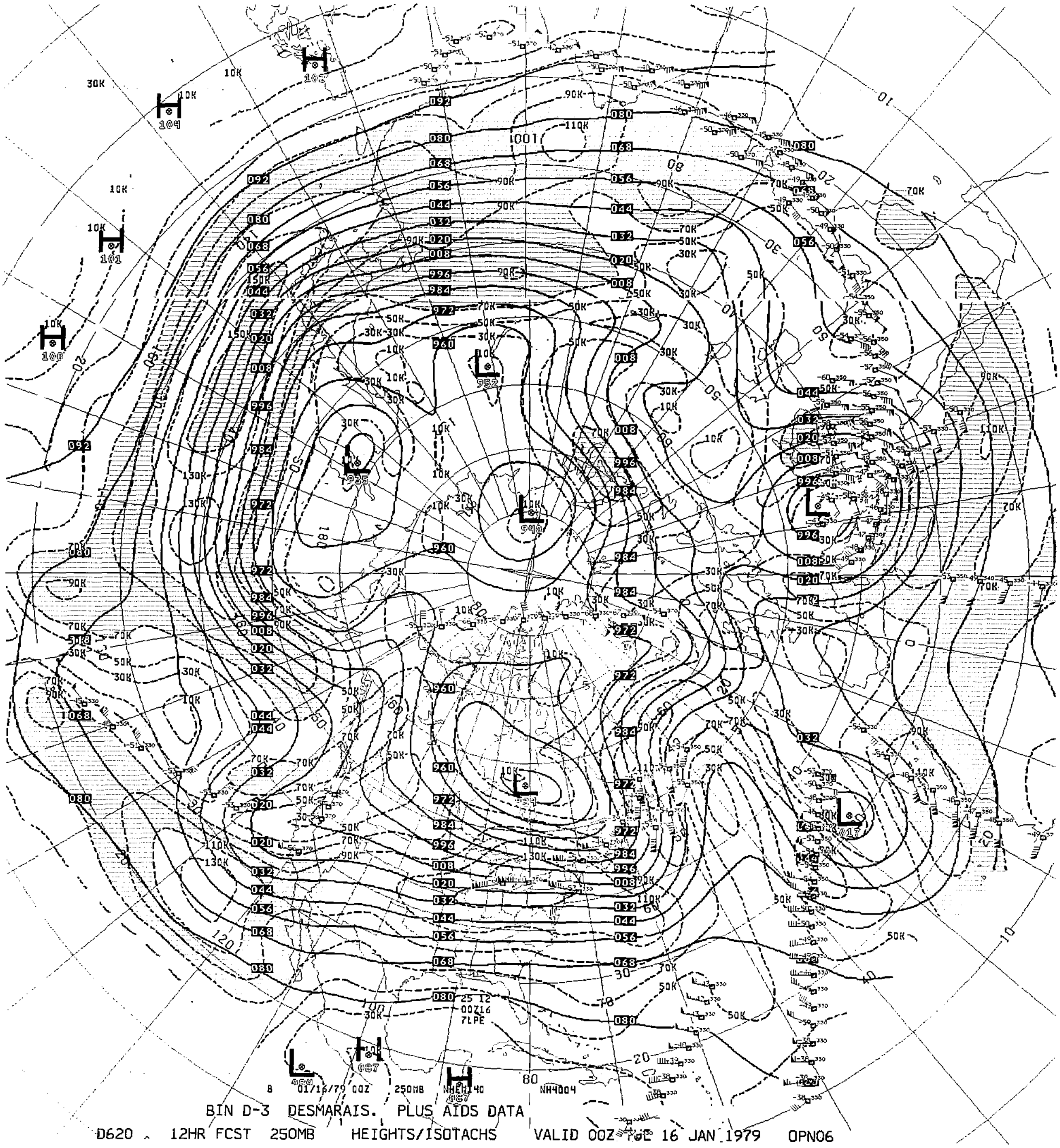


Figure 15. Data from the Aircraft Integrated Data System, similar to ASDAR, for 0000 GMT 16 January 1979.

Figures 13 and 14 illustrate the typical coverage by aircraft at all levels in the Northern and Southern Hemisphere, respectively. Most conventional aircraft reports are taken at checkpoints along the flight path; these are usually at 5° longitude intervals. Consequently, the aircraft reports tend to appear along meridians. In a six-hour interval, a busy route may produce 20 or more observations within a kilometer or two in the vertical and a few degrees in latitude, all on the same meridian.

An exception to this practice occurs with the ASDAR device referred to earlier in this lecture. ASDAR is basically a communications device which periodically interrogates the on board computers of the aircraft's inertial navigation system, and transmits selected information to ground readout stations by geostationary satellite. That information includes aircraft position and flight level, the time, the wind, and the temperature. During the Global Weather Experiment, more than a dozen wide-bodied aircraft were equipped with ASDAR devices adjusted to report 8 times each hour. Thus, ASDAR reports appear on the map as a nearly continuous stream marking the flight path of the aircraft. Experience has suggested that the ASDAR winds are extremely accurate and reliable. Indeed, there is some suspicion that ASDAR winds at high levels are more accurate than radiosonde winds. Figure 15 illustrates a typical set of ASDAR data on a commercial flight from Rome to Hong Kong across southern Asia in January 1979. Note the high degree of internal consistency in the data, and the qualitative agreement with the superimposed analysis (which did not have access to this particular set of data).

It is expected that ASDAR units will be installed on many of the world's fleet of commercial jet aircraft by 1985. Meteorologists can happily anticipate an abundance of highly accurate wind data, at least at one level, by that time.

IV. Data Processing and Preliminary Quality Control*

Remote temperature soundings and cloud-motion wind vectors from the two U.S. geostationary satellites are processed by the U.S. National Environmental Satellite Service (NESS). After processing, selected data sets are transmitted to users over the Global Telecommunications System (GTS). However, because NESS and NMC share computing facilities, these data are made available to NMC in common disk storage. No further processing is performed by NMC on this data prior to its ingestion into the global data assimilation system.

Data from radiosondes and aircraft generally arrive at NMC via teletypewriter circuits. These data are subjected to various kinds of processing to render them suitable for use in the assimilation system. The incoming traffic initially is enqueued in a disk storage device. This queue is repetitiously scanned by a program which examines the bulletin headings to determine the type of data involved, and then directs the bulletins to one of several queues where they await further processing by an appropriate decoder. The decoders are called upon

---

* Material in this section taken largely from McDonell (1973, 1975).

repeatedly during the day in order to accumulate those reports which form the parts of the observational data base. Priority is given to the running of the decoders at certain critical periods to insure that useful reports needed for operational requirements are not left stranded in the disk storage queues.

Since bulletin heading recognition is critical to the system, there are provisions for the unrecognizables to be displayed at a bulletin heading quality control entry and display station where they are examined by communications technicians who make corrections if possible. Corrected bulletins are resubmitted to the system as incoming traffic. There should be no important bulletin which the system refuses, and any such appearance at the display station leads to prompt action to add the bulletin to the list of legitimates.

In the case of the radiosonde decoder, any discrepancies in the identification groups of the message cause the report to be enqueued in another disk area for inspection at a data quality control entry and display station manned by meteorological technicians who make corrections and re-submit the report to the decoder. Using this same facility, upper-air data from land and ship stations can be examined to determine the contents of the report after processing. While a report is visible on the display device, corrections or other entries can be made after which the report is reinserted back into the processed data base.

Considerable bookkeeping is involved in filing the reports in the disk storage. For example, the station's international index number is matched against a master list of possible numbers in order to determine its existence and to append the location, elevation, and receipt time to the report. If a match is found, the report is stored and a flag is set to indicate this result. If a redundant report is encountered at a later time, it is by-passed based on a simple test of this flag.

Aircraft reports are treated with bookkeeping procedures considerably different from those for land stations because their locations, elevation, and times are not predetermined by a schedule. They are decoded and enqueued in their order of receipt. As a result of a routine which sorts the reports according to time, location, and flight level, the duplicates end up side-by-side. A pass through this array is then made keeping only the non-redundant reports. Similar procedures also are required for reports from ships, cloud-motion winds, and drifting buoys.

As a result of the report processing described above, the observational data reside as a set of files in the disk storage device with contents as follows:

1. Surface reports from land stations.

2. Surface reports from ships and drifting buoys.

3. Manually prepared surface reports (monitoring bogus and estimates of relative humidity).

4. Upper-air reports from land stations, ships, and reconnaissance aircraft (dropsondes and single-level data).

5. Aircraft reports (except for reconnaissance).

6. Cloud-motion wind vectors.

7. Upper-air reports from TIROS-N and NOAA-6.

The reports in the fourth file are examined and checked for vertical consistency provided adequate information is available. For height and temperature data, the testing utilizes the hydrostatic equation and "flags" parameters to indicate the results of the test. For winds the testing consists of examining the wind sounding for unlikely vertical changes in direction and/or speed. Consideration has been given to the reporting procedures in designing these tests. For example, a common error when encoding the temperature group in the radiosonde code is to make the tenths of temperature an even digit (indicating a positive value) when the temperature is in fact negative, or as an odd digit (indicating a negative value) when the temperature is in fact positive. In recomputing erroneous temperatures a test is made to determine if one of these conditions caused the trouble and if so the reported value with the opposite sign is substituted. By using the reported maximum wind and the tropopause wind, the wind testing program validates nearby mandatory level winds before proceeding to test the other levels. In order to keep a record of the results of these tests a system of "flagging" the data is used in which the flags are appended to the parameters to indicate the result of the testing. These flags are printed out when the data are listed and are available to other programs which later use the data.

We have found it convenient to prepare the data for the particular analysis that will use it. To achieve this the data are regrouped and reformatted to make efficient use of the word size of the computer and the disk storage devices. In addition the reports are condensed to retain only those parameters that are to be used by the analysis. As has been noted in an earlier lecture, three separate analysis systems are presently used at NMC. Each has its own pre-processor, although many of the elements of each pre-processor are common to all three. As an example, the preprocessing steps described below are those which serve the successive-corrections analysis supporting the limited-area fine mesh (LFM) prediction model.

The first six files[1] are regrouped into (1) a file of surface data and (2) a file of upper-air data, keeping only the parameters needed by the analysis programs. Some of the parameters are merely extracted and others require computations involving the entire report. The information that is common to all reports in both files are the report's identification, time, type, geographical location, and elevation.

[1] Remote sounding data are not presently used in the limited area analysis because of its short data cutoff time.

For each report in the surface file, the sea level pressure, surface temperature, and surface wind are extracted, and the surface relative humidity is computed from the temperature and dewpoint. Estimates of the mean relative humidity in each of three layers between the surface and approximately 500 mb are made by interrogating the current weather and cloud information. The surface relative humidity contributes only to the estimate for the lowest layer (a 50 mb thick layer above the surface).

For each report in the upper-air file, the height, temperature, and wind for the mandatory levels (1000 to 100 mb) as well as the pressure and temperature for the tropopause level(s) are extracted. The following additional steps are also performed at this time:

(1) Forecast wind shears are computed between the mandatory pressure levels and between the tropopause levels in the two mandatory levels bounding the tropopause level at the location of each aircraft. The appropriate forecast shears between the aircraft level and the mandatory levels are vectorially added to the aircraft wind report and the results inserted as the mandatory level wind values to be used in the analysis. By examining the pressure altitude of the aircraft, the programs can determine the "off-level" distance between the report and the mandatory level being analyzed.

(2) The mean relative humidity for three layers between the surface and approximately 500 mb are computed using the mandatory level, and if available, the significant level temperature and dewpoint depression values. The pressure values bounding these layers are determined by using the reported station pressure and the forecast tropopause pressure.

(3) The reported 850 and 700 mb heights and the 700 mb temperature are used to calculate a "free-air" 850 mb temperature for each station whose elevation exceeds 1200 meters, resembling closely the method used at NMC when the analyses were done manually. This "free-air" 850 mb temperature is computed by assuming that temperature varies linearly with the logarithm of pressure and ignoring the difference between virtual and actual temperature, i.e., ignoring the moisture. For height (Z) in meters and temperature (T) in degrees Celsius, this relationship is expressed as follows:

$$T_{850} = .352\ (Z_{700} - Z_{850}) - T_{700} - 546.3$$

where the subscripts are pressure in millibars.

(4) Corrections for the solar radiation effects on the thermistor element are applied to the mandatory level height and temperature data above the 150 mb level.

(5) The flags from the vertical consistency checks described earlier are translated to numerical values and retained with the parameters.

Following completion of these steps, the data are considered suitable for the analysis procedure.

References

Bauer, K., 1976: "A comparison of cloud motion winds with coinciding radiosonde winds." Mon. Wea. Rev., 104, 922-931.

Bergman, K., 1978: "Role of observational errors in optimum-interpolation analysis." Bull. Amer. Meteor. Soc., 59, 1603-1611.

Bruce, R. L., Duncan, and J. Pierluissi, 1977: " Experimental study of the relationship between radiosonde temperatures and satellite-derived temperatures." Mon. Wea. Rev., 105, 494-496.

Fleming, R., T. Kaneshige, and W. McGovern, 1979: "The Global Weather Experiment, 1. The observational phase through the first Special Observing Period." Bull., Amer., Meteor. Soc., 60, 649-659.

Gemmill, W., and S. Larson, 1979: "Real time ocean thermal structure analysis." Proc. IOC/WMO Seminar on Oceanographic Products and the IGOSS Data Processing and Services System, Moscow, 2-6 April 1979. Intergovernmental Oceanographic Commission Report No. 17.

McDonell, J., 1973: Notes on operational objective analysis procedures. Unpublished NMC manuscript prepared for NWS Advanced Prediction Techniques Course, September 1973.

______________, 1975. "Preparation of the data base for use in an operational environment." Proc. JOC Study Group Conf. on Four-Dimensional Data Assimilation, Paris, 17-21 Nov. 1975. GARP Working Group on Numerical Experimentation Report No. 11, January 1976.

Morone, L., 1979:. Revised observational error statistics for TIROS-N A and C retrievals and cloud-tracked winds. NMC memorandum for the record, May 17, 1979.

Phillips, N., L. McMillin, A. Gruber, and D. Wark, 1979. "An evaluation of early operational temperature soundings from TIROS-N." Bull. Amer. Meteor. Soc., 60, 1188-1197.

Phillips, N., 1980. "Cloudy winter satellite retrievals over the extratropical Northern Hemisphere oceans." NMC manuscript to be submitted for publication.

Schlatter, T., 1979. "An assessment of operational TIROS-N retrievals over the United States." Submitted to Mon. Wea. Rev.

Seaman, R., 1977. "Absolute and differential accuracy of analyses achievable with specified observational network characteristics." Mon. Wea. Rev., 105, 1211-1222.

Tracton, M., A. Desmarais, R. van Haaren, and R. McPherson, 1980. "The impact of satellite soundings upon the National Meteorological Center's analysis and forecast system - the data systems test results." To be published in Mon. Wea. Rev., April 1980.

Wark, D., and H. Fleming, 1966. "Indirect measurements of atmospheric temperature profiles from satellites: I. Introduction." Mon. Wea. Rev., 94, 351-362.